Converting Operation Database to Star Schema Structure

A Thesis submitted to Faculty of Information Technology in partial fulfillment of the requirements for the degree Master

(Information Technology),

University Utara Malaysia

By

Mohammed Aal Mohammed (801602)

Syahida Binti Hassan

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

MOHAMMED AAL MOHAMMED
(801602)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**

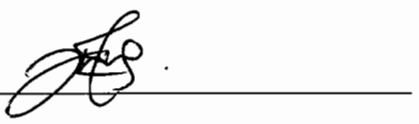
telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

CONVERT OPERATION DATABASE TO STAR SCHEMA STRUCTURE

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **MISS SYAHIDA HASSAN** syahida hassan

Tandatangan
(Signature) : 

Tarikh
(Date) : 20 May 2010

PERMISSION TO USE

In presenting this project of the requirements for a Master of Science in Information and Communication Technology (MSc. IT) from Universiti Utara Malaysia, I agree that the University library may make it freely available for inspection. I further agree that permission for copying of this project paper in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or in their absence, by the Dean of Graduate School. It is understood that any copying or publication or use of this project or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from my project paper.

Request for permission to copy or make other use of materials in this project, in whole or in part, should be addressed to:

Dean of Graduate School

University Utara Malaysia

06010 Sintok

Kedah Darul Aman

Malaysia

ABSTRACT

Today's, structural data representation has been developed rapidly in term of complexity. Different studies addressed the advantages of employing five star schemas in simplifying and managing the data components based techniques. Moreover, other researchers highlighted the importance of using data warehouse in enhancing the database abilities and handling large amounts of data in different structural representation. Hence, this study aimed to apply the five star schema structures in simplifying the data retrieval process. System development research methodology has been employed in this study for analyzing and developing the proposed system. Result has been conducted based on the using of use test case technique.

ACKNOWLEDGEMENTS

Praise to Allah for his guidance and blessing for giving me the strength and perseverance to complete this project. I would foremost like to thank my parents, for providing me with the opportunity to pursue my goals and for their love and affection, which has helped me through the most trying times. Equal gratitude goes out to my siblings and brothers. I would like to thank my supervisor: Syahida Binti Hassan for her guidance and constant that has enabled me to complete my project work. Moreover, I would also like to thank her for the opportunities that she has made available to me.

Furthermore I like to thank all my sisters and brothers for their kindnesses and supports, as well as all lectures in the faculty of Information Technology, that they gave to me all of the information and knowledge for completing my study. "May Allah bless all of them".

Mohammed Aal Mohammed / 2010-05-15

TABLE OF CONTENTS

CHAPTER FOUR

RESULT ANALYSIS

CHAPTER FIVE

EVALUATION

CHAPTER SIX

CONCLUSION

LIST OF FIGURES

LIST OF TABLES **Page No.**

CHAPTER ONE

INTRODUCTION

1. Introduction

Decision Support Systems (DSS) are rapidly becoming essential to achieving a competitive advantage for businesses. DSS allows businesses to obtain a huge amount of data that are locked in operational databases and other sources of data and turn data into useful information. Many companies have already construct or are constructing decision-support databases called data warehouses, in which users can perform their data analysis. A typical data warehouse extracts, integrates and manages the relevant information from multiple, independent, heterogeneous data sources into one centralized repository of information to support decision-making needs of knowledge workers and decision makers in the form of Online Analytical Processing (OLAP) (Han and Kamber 2001; Radiant, V., 2007).

The contents of the thesis is for internal user only

References

Adamson, C., (2006). Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance. Relative on July 19, 2006.

Aisha, L., Younus, J., & Ayesha, A. (2009). Hybrid Technique of using ANN in Semi-Star Schema Generation. International Journal of Hybrid Information Technology Vol.2, No.2, April, 2009.

Alberto, A., & Carme, M. (2001). A Bitemporal Storage Structure for a Corporate DataWarehouse. The Spanish Research Program PRONTIC under project TIC2000-1723- C02-01.

Angela, B., Fabiano, C., & Stefano, C. (2001). Designing Data Marts for Data Warehouses. ACM Transactions on Software Engineering and Methodology, Vol. 10, No. 4, October 2001, Pages 452–483.

Behrooz, S. (2003).Teaching Effective Methodologies to Design a Data Warehouse. Retrieved on 25 March 2010, from (http://proc.isecon.org/2001/35c/ISECON.2001.Seyed-Abbassi.pdf).

Bennett, S., McRobb, S., & farmer, R. (2002). Object-oriented System Analysis and Design 2 Edition. UK, McGraw Hill.

Booch, G., Rumbaugh, J., & Jacobson, I. (2001).Unified Modeling Language Semantics and Notation Guide1.0. San Jose, CA: Rational Software Corporation). Retrieved on 17 Sep 2009, from (www.cs.wustl.edu/~kjg/cse132/.../UML%20notation%20guide.pdf).

Breslin, M. (2004). Data warehousing battle of the giants: Comparing the basics of the Kimball and Inmon Models. Business Intelligence Journal (Winter 2004) 6-20.

Charles, J., and Grry, P. (2008). Oracle Database Security: Preventing Enterprise Data Leaks a t the Source. Oracle Corporation.

Chen, X., Neil, P., & Neil, E. (2008). Adjoined Dimension Column Clustering to Improve Data Warehouse Query Performance. ICDE 2008, pp. 1409-1411.

Connolly, T., & begg, C. (2010). A Practical Approach to Design, Implementation, and Management fifth edition .

Connolly, T & Begg, C. (2004). Database Solutions. A step-by-step guide to building databases. Related from www.booksites.net.

Connolly, T & Begg, C. (2000).Database Systems - A Practical Approach to Design, Implementation and Management second edition .

Daniel, J., Goh, N., & Yusop, S. (2007). Data Transformation Services (DTS): Creating Data Mart by Consolidating Multi-Source Enterprise Operational Data. World Academy of Science, Engineering and Technology 34 2007.

David, S., Karen, C., & Ozgur, T. (2005). Comparing the Effect of Alternative Data Warehouse Schemas on End User Comprehension Level. Retrieved on 27 March 2010, from (http://mis.temple.edu/sigdss/icis05/proceedings/dssworkshop05-schuff.pdf).

Dennis, A., Wixom, B.H., & Tegarden, D. (2005). System analysis and design with UML version 2.0: an object-oriented approach with UML, 2nd edition. Hoboken, NJ: John Wiley and Sons, Inc.

Gary, P., and Greg, W. (2006). What Would an Exemplary Entrepreneurship Dataset Look Like? Imperatives and Opportunities for Research, October 26-27, 2006, Washington, DC.

Hasso, P. (2009). A Common Database Approach for OLTP and OLAP Using an In-Memory Column Database. SIGMOD'09, June 29–July 2, 2009, Providence, Rhode Island, USA. ACM 978-1-60558-551-2/09/06.

Hoffer, J. A., George, J. F & Valacich, J. S. (2002). Modern Systems Analysis and Design (3rd Edition). Upper Saddle River, New Jersey: Prentice Hall.

Heise, D. (2007). EDUCAUSE Decision Support Data Warehouse Constituent Group. Downloaded from dheise.andrews.edu/dw/DWData.html.

Husemann, B., Lechtenborger, J., & Vossen, G. (2000). Conceptual data warehouse design. In Proceedings of International Workshop on Design and Management of Data Warehouses, Stockholm, 2000.

Inmon, H. (2002). Building the Data Warehouse, 3rd Edition. John Wiley & Sons, Inc., New York, NY, USA, 2002

Jacobson, I., Christerson, M., Johnsson, P. & Overgaars, G. (2004). Object-oriented Software Engineering: A Use Case Driven Approach (revised).

Kang (2002). Exploiting Versions for On-Line Data Warehouse Maintenance in MOLAP Servers, Proc. Of VLDB Conference, China 2002.

Khurram, S. (2008). Semi-star Schema for Operational and Analytical Requirements of SMEs. International Journal of Management and Decision Making (IJMDM): Special Issue on "Decision Support System and Knowledge Management in SME's", Greece, 2008.

Kimball, R. & Ross, M. (2002). The Data Warehouse Toolkit, Second Edition, Wiley, 2002.

Kimball, R., & Mertz, R. (2000). The Data Web house Toolkit: Building the Web-Enabled Data Warehouse. John Wiley & Sons, Chichester, 2000.

Latif, A. Javed, Y., & Khan, S. (2008). Semi-automated approach for converting ERD to semi-star schema. Emerging Technologies, 2008. ICET 2008. 4th International Conference on. 978-1-4244-2210-4.

Levene, M., & Loizou, G. (2001). Guaranteeing no interaction between functional dependencies and tree-like inclusion dependencies. Theoretical Computer Science, 254:683–690, 2001.

Liedes, A., and Wolski, A. (2006). SIREN: A memory-conserving, snapshot-consistent checkpoint algorithm for in-memory databases. In ICDE '06, page 99, 2006.

Marcin, G. (2009). Extended Cascaded Star Schema and ECOLAP Operations for Spatial Data Warehouse. Intelligent Data Engineering and Automated Learning - IDEAL 2009.

Mark, L., & George, L. (2003). Why is the Snowflake Schema a Good Data Warehouse Design? Retrieved on 27 March 2010, from (http://www.dcs.bbk.ac.uk/~mark/download/star.pdf).

Marotta, A., & Ruggia, R. (2002). Data Warehousing Design: A Schema-transformation Approach. In proceedings of 22nd International Conference of the Chilean

Computer Science Society, 2002. SCCC 2002, IEEE Computer Society, Atacama, Chile, 6-8 November 2002, pp 153- 161.

Mendelzon (2000). Temporal Queries in OLAP, Proceedings of VLDB Conference, Egypt, 2000.

Nancy, R., & Cary, N. (2005). Stars and Models: How to Build and Maintain Star Schemas Using SAS Data Integration Server in SAS. SAS Institute Inc.

Pasha, A., Nasir, A., and Shahzad, K. (2004). Semi-Star Modeling Schema for Managing Data Warehouse Consistency. National Conference on Emerging Technologies 2004.

Paplpanas, T. (2000). Knowledge Discovery in Data Warehouses," ACM SIGMOD Record, 29(3), pp. 88-100.

Pasha, A., Nasir, A., & Shahzad, K. (2004). Semi-Star Modeling Schema for Managing Data Warehouse Consistency. National Conference on Emerging Technologies 2004.

Red Brick Systems. Star Schemas and STARjoin Technology. Los Gatos, CA: Red Brick Systems (1995). Retrieved on 28 March 2010, from (www.redbrick.com).

Rifaie, M., Blas, E., Muhsen, A., Mok, T., Kianmehr, K., Alhajj, R., & Ridley, M. (2008). Data Warehouse Architecture For GIS Applications.

Radiant, V. (2007).Decision Support System Architecture, Hardware, and Operating System Platforms. Jurnal Sistem Informasi Vol. 2 No. 1.

Rob, P., & Carlos, C. (2000). Database Systems: Design, Implementation, and Management, Course Technology, Cambridge.

Santhi, K., and Jigeesh, N. (2010). A Virtual Data Warehouse for Manufacturing Industry. Social science research network 2010.

Shi, H., Tung, C., and Jia, S. (2007). Data warehouse enhancement: A semantic cube model approach. Information Sciences 177 (2007) 2238– 2254.

Schmuller, J. (2002). SAMS teach yourself UML in Hours. SAMS Publishing, Indiana.

Seyed-Abbassi, B. (2002). Teaching Effective Methodologies to Design a Data Warehouse.

Shi, H., Tung, C., & Jia, S. (2007). Data warehouse enhancement: A semantic cube model approach. Information Sciences 177 (2007) 2238– 2254.Systems (1995). Retrieved on 28 March 2010, from (www.redbrick.com).

Stonebraker, M., and Cetintemel, U. (2005). One size fits all": An idea whose time has come and gone. In ICDE '05, pages 2–11, 2005.

Tim, M. (2004). Reconsidering Multi-Dimensional Schemas. SIGMOD Record, Vol. 33, No. 1, March 2004.

Van, L. (2008). A Data Warehouse Model for Micro-Level Decision Making in Higher Education." The Electronic Journal of e-Learning Volume 6 Issue 3 2008, pp. 235 – 244, available online at www.ejel.org.

Yeol, S., Ritu, K., Yuan, A., Suan, L., Sang, P., Jinho, K., & Yang, M. (2008). SAMSTAR: An Automatic Tool for Generating Star Schemas from an Entity-Relationship Diagram. Conceptual Modeling - ER 2008.

Yingying, T., Qiang, Z., Calisto, Z., & Wing, L. (2008). Optimizing Large Star-Schema Queries with Snowakes via Heuristic-Based Query Rewriting. IBM Canada Ltd 2008.